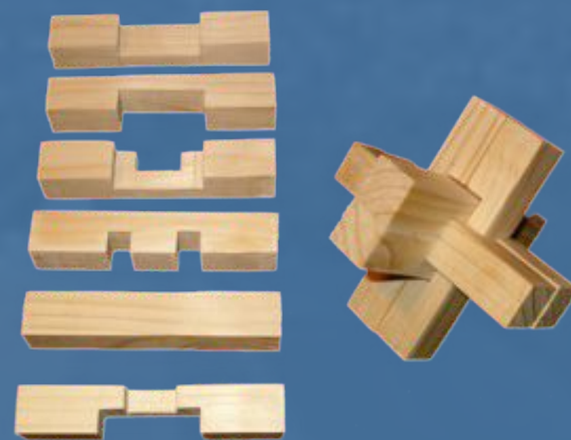
The Challenges Ahead



Filipe Maia
Petascale Postdoctoral Fellow
NERSC, Lawrence Berkeley National Lab

2012-02-17

Thanks to

UPPSALA UNIVERSITET

Tomas Ekeberg

Max Hantke

DESY

Anton Barty

Andrew Martin

SLAC NATIONAL ACCELERATOR LABORATORY

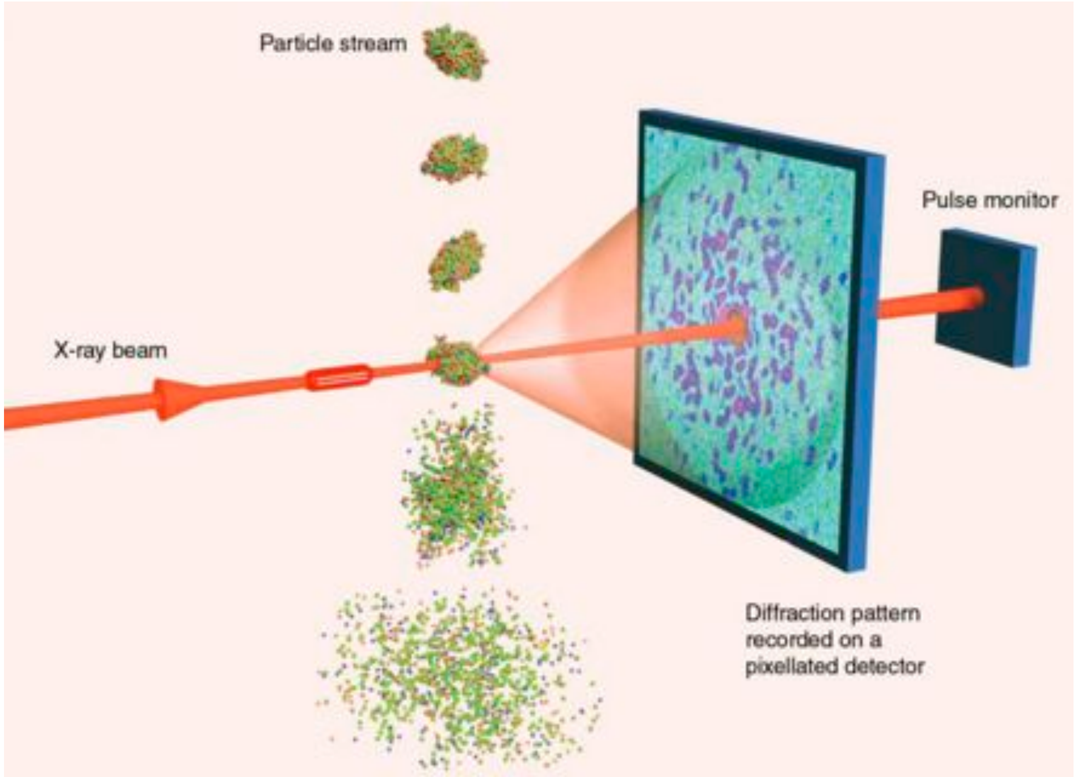
Duane Loh

BERKELEY LAB
Lawrence Berkeley National Laboratory

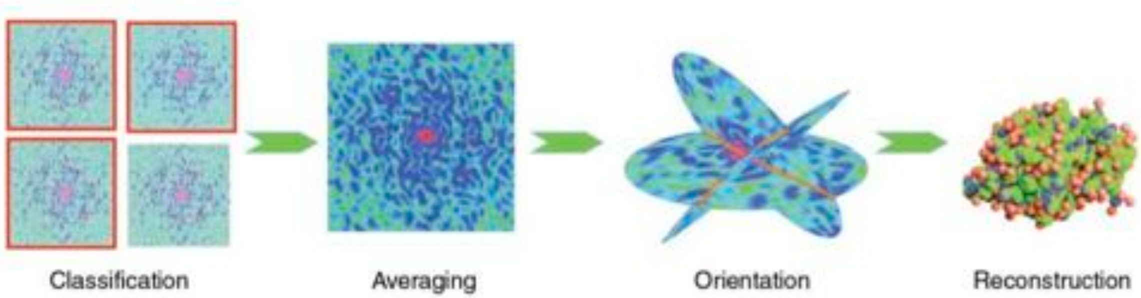
Stefano Marchesini

Chao Yang

A typical X-ray coherent diffractive imaging experiment



1. An X-ray beam hits the sample
2. It's scattering pattern is recorded
3. The structure of the sample is recovered from the pattern.

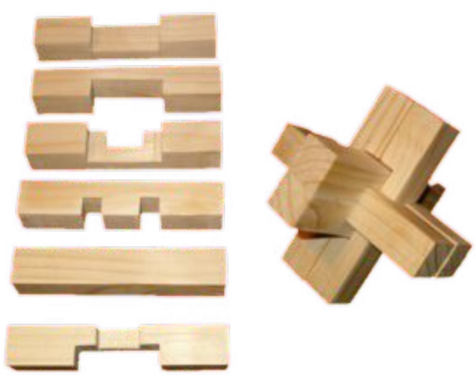


Gaffney, K.J. & Chapman, Science, 316, 1444-8 (2007).

What are the major challenges ahead for X-ray Coherent Diffraction Imaging?

Data Deluge

Validation



3D Assembly

Data Deluge

The number of X-ray free electron lasers is rapidly increasing.

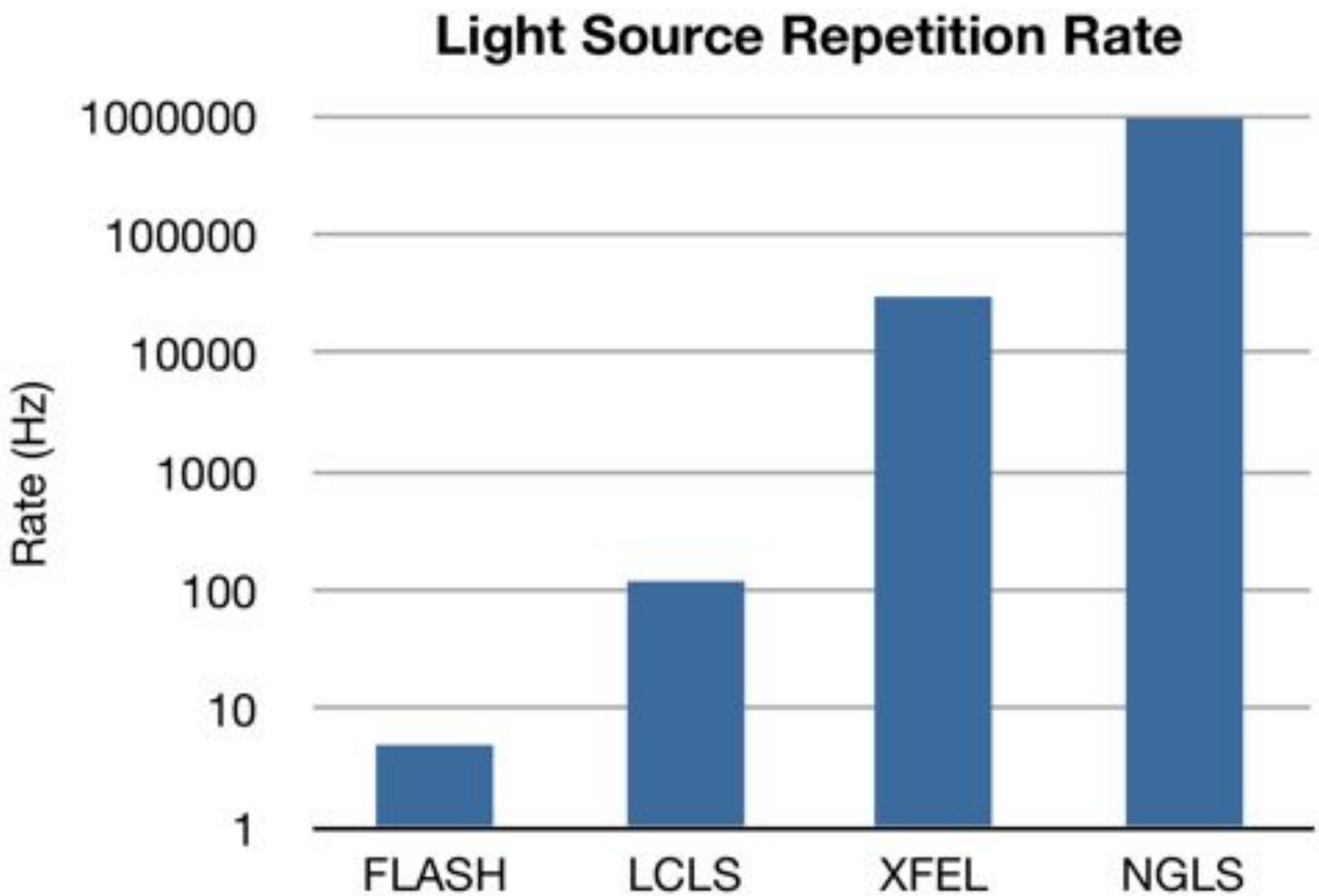
Table 2. List of SASE X-ray Free-Electron Laser User Facilities Currently Proposed, in Construction, or in Operation [a]

Facility	Proposed By	Wavelength Range	Status	Operation
Arc-en-Ciel	Synchrotron SOLEIL, France	200 nm–0.2 nm	Proposal	
European XFEL	XFEL Company, Germany	2 nm–0.1 nm	Construction	2015
FERMI @ Elettra	Elettra, Italy	80 nm–4 nm	Commissioning	2011
FLASH	DESY, Germany	60 nm–5 nm	Operation	2005
JLAMP	Jefferson Laboratory, USA	120 nm–10 nm	Proposal	2016
LCLS	SLAC National Accelerator Lab, USA	2.5 nm–0.12 nm	Operation	2009
MAX-IV	MAXLab, Sweden	to 1 nm	Proposal	
Next Gen. Light Source	Lawrence Berkeley National Lab, USA	to 1.2 nm	Proposal	2020
Pohang XFEL	Pohang Light Source, Republic of Korea	to 0.1 nm	Proposal	2015
Polish FEL	Soltan Inst. for Nuclear Studies, Poland	to 9 nm	Proposal	2015
SCSS	SPring-8 Light Source, Japan	to 0.1 nm	Construction	2010
SCSS Test Accelerator	SPring-8 Light Source, Japan	to 50 nm	Operation	2006
Shanghai FEL	Shanghai Inst. of Applied Phys., China	to 9 nm	Construction	2015
SPARX	University of Rome Tor Vergata, Italy	40 nm–0.5 nm	Proposal	2013
Swiss XFEL	Paul Scherer Institute, Switzerland	to 0.1 nm	Proposal	2016
TAC SASE FEL	Turkish Accelerator Complex, Turkey	to 7 nm	Proposal	

[a] The quoted wavelength ranges are, in some instances, reached with post-construction upgrades.

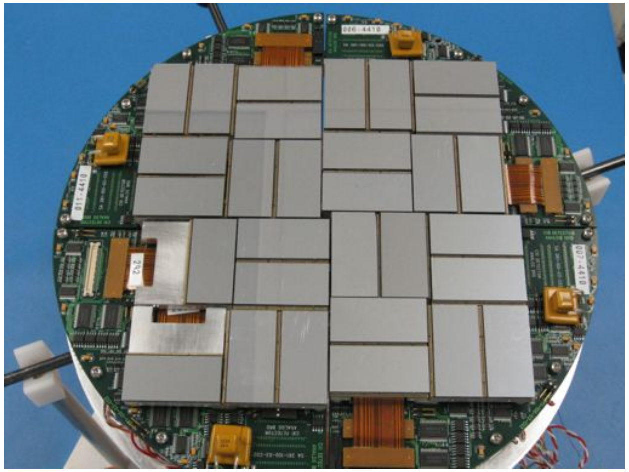
Galayda, J.N. et al. X-ray free-electron lasers - present and future capabilities. *JOSA B* 27, B106-B118 (2010).

The repetition rate of light sources if growing exponentially …

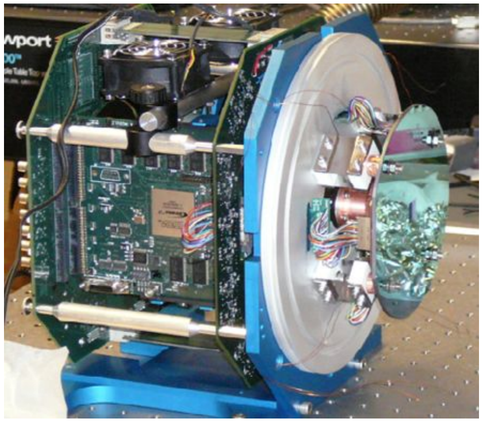


… and the detectors are doing their best to keep up with it.

Pilatus 1M,
frame rate 30 Hz

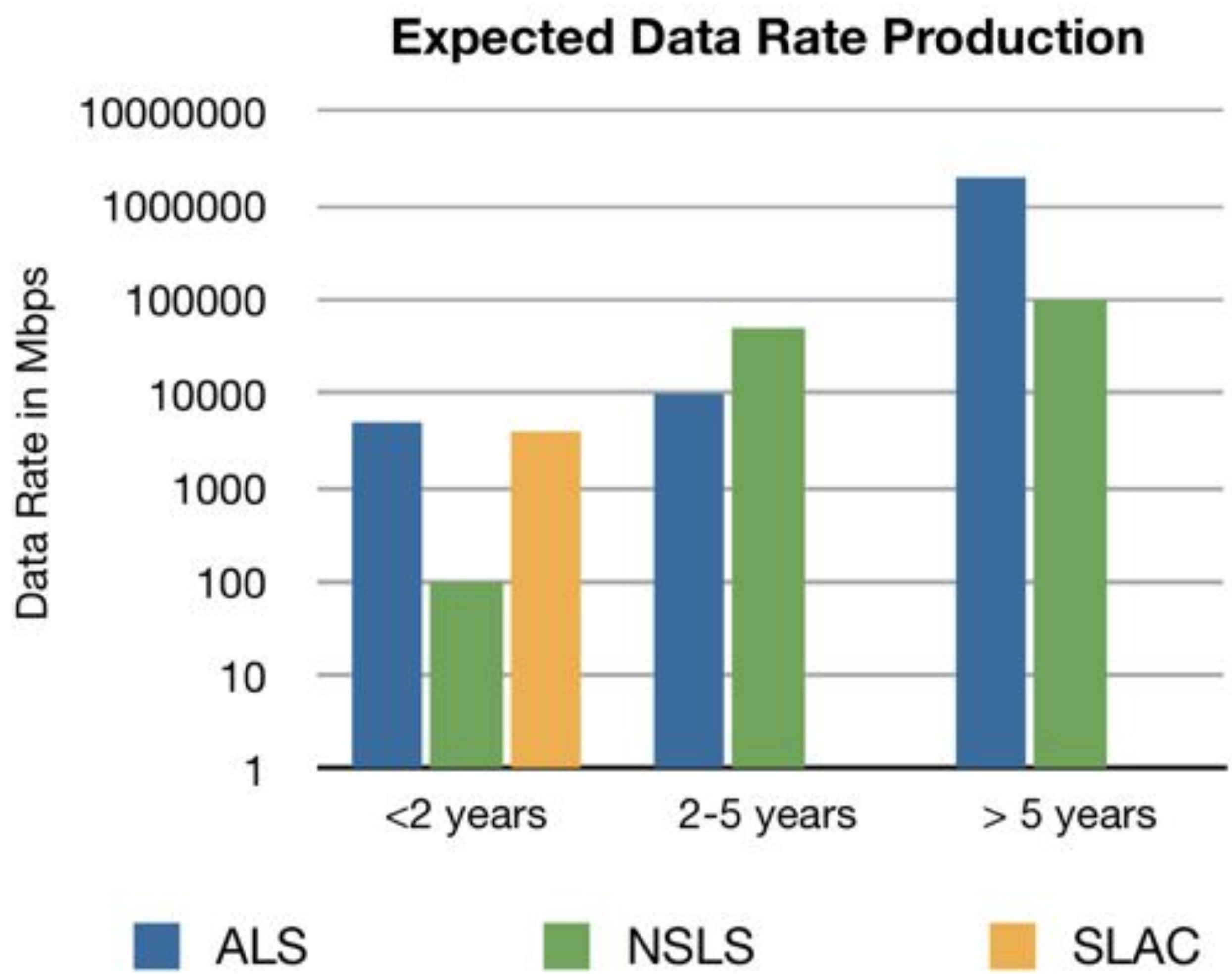


CXI detector,
frame rate 120 Hz



FastCCD,
frame rate 200 Hz

The growth in data rate is not confined to free electron lasers.



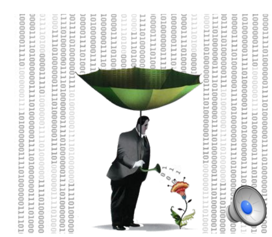
Report of the Basic Energy Sciences Network Requirements Workshop - 2010

Data logistics

European XFEL

100 GB/sec
370 TB/hr

Data logistics



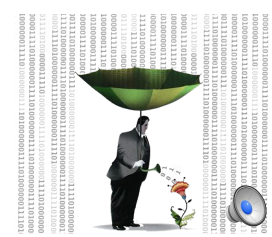
Data logistics

We are not the only ones with this problem

"Most scientific disciplines are finding the data deluge to be extremely challenging, and tremendous opportunities can be realized if we can better organize and access the data."

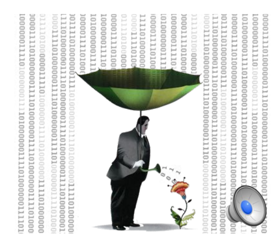
-- *Science* Staff

Special Issue:
Dealing with Data, *Science*, February 2011



Sorting the raw data in categories for later analysis is challenging

- Millions of images to go through
- Often low signal to noise ratios
- Many approaches yet to be explored



Sorting using machine learning algorithms

Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering

Chun Hong Yoon,[1] Peter Schwander,[1] Chantal Abergel,[2] Inger Andersson,[3]
Jakob Andreasson,[4] Andrew Aquila,[5] Saša Bajt,[5] Miriam Barthelmess,[5] Anton Barty,[6]
Michael J. Bogan,[7] Christoph Bostedt,[8] John Bozek,[8] Henry N. Chapman,[6,9]
Jean-Michel Claverie,[2] Nicola Coppola,[10] Daniel P. DePonte,[6] Tomas Ekeberg,[3]

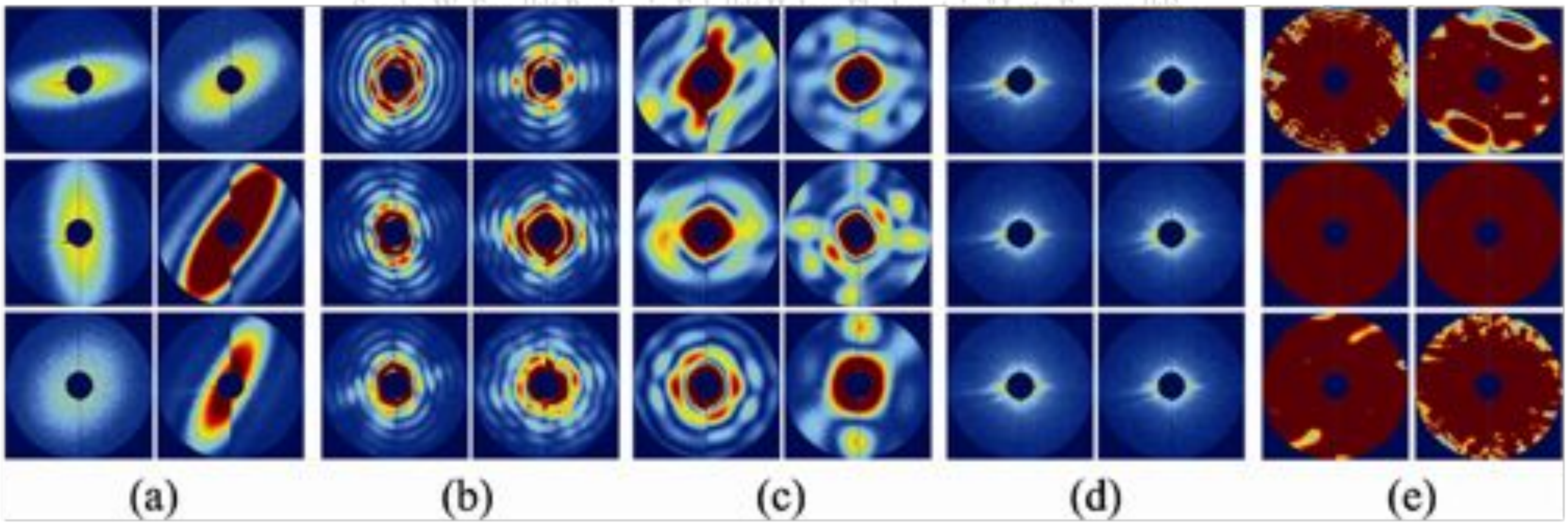
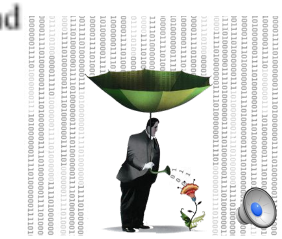
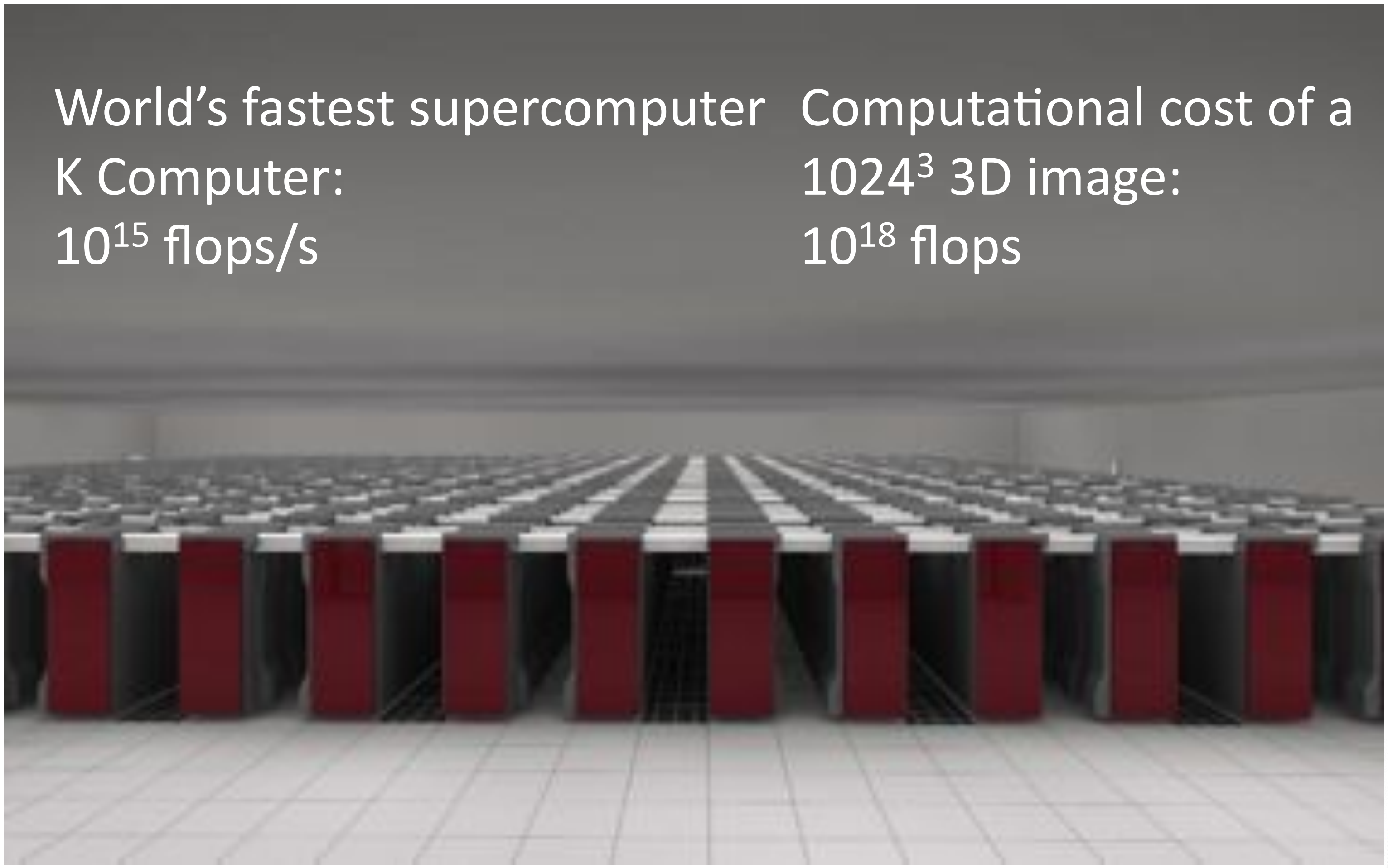


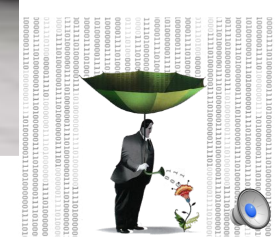
Fig. 2. Randomly selected representatives from (a) Cluster 2: nanorice snapshots; (b) Cluster 7: mimivirus snapshots; (c) Cluster 10: miscellaneous snapshots; (d) Class 9: blank snapshots; and (e) Cluster 14: saturated snapshots.



Yoon, C.H. et al. Optics Express 19, 16542-9 (2011).

The computational requirements will grow such that making scalable software will be a necessity





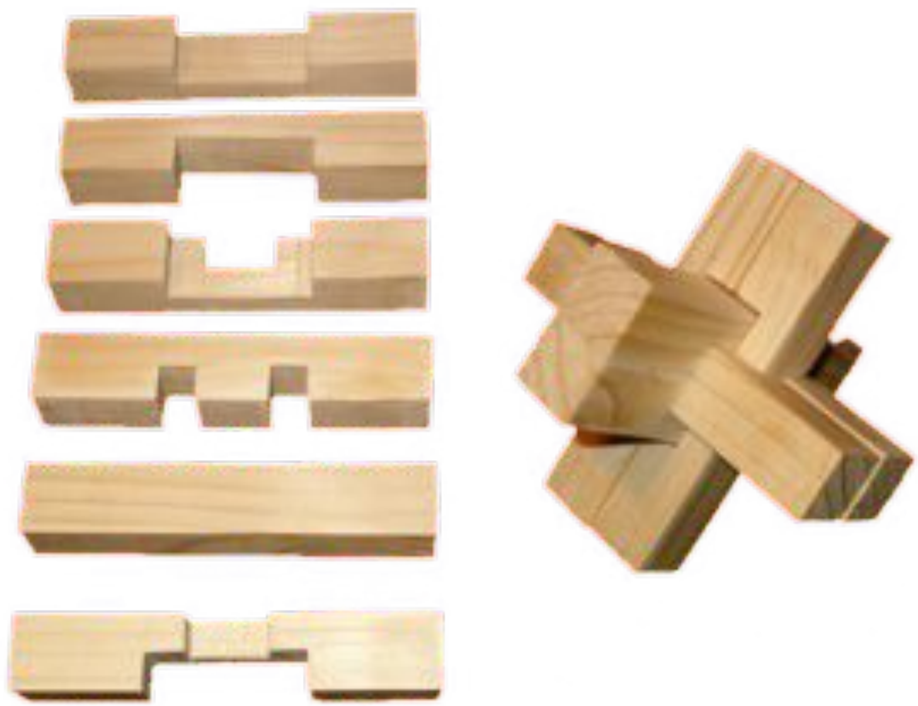
Data sharing and standard file formats are necessary building blocks for improved analysis software

RCSB PDB
PROTEIN DATA BANK

CCP4

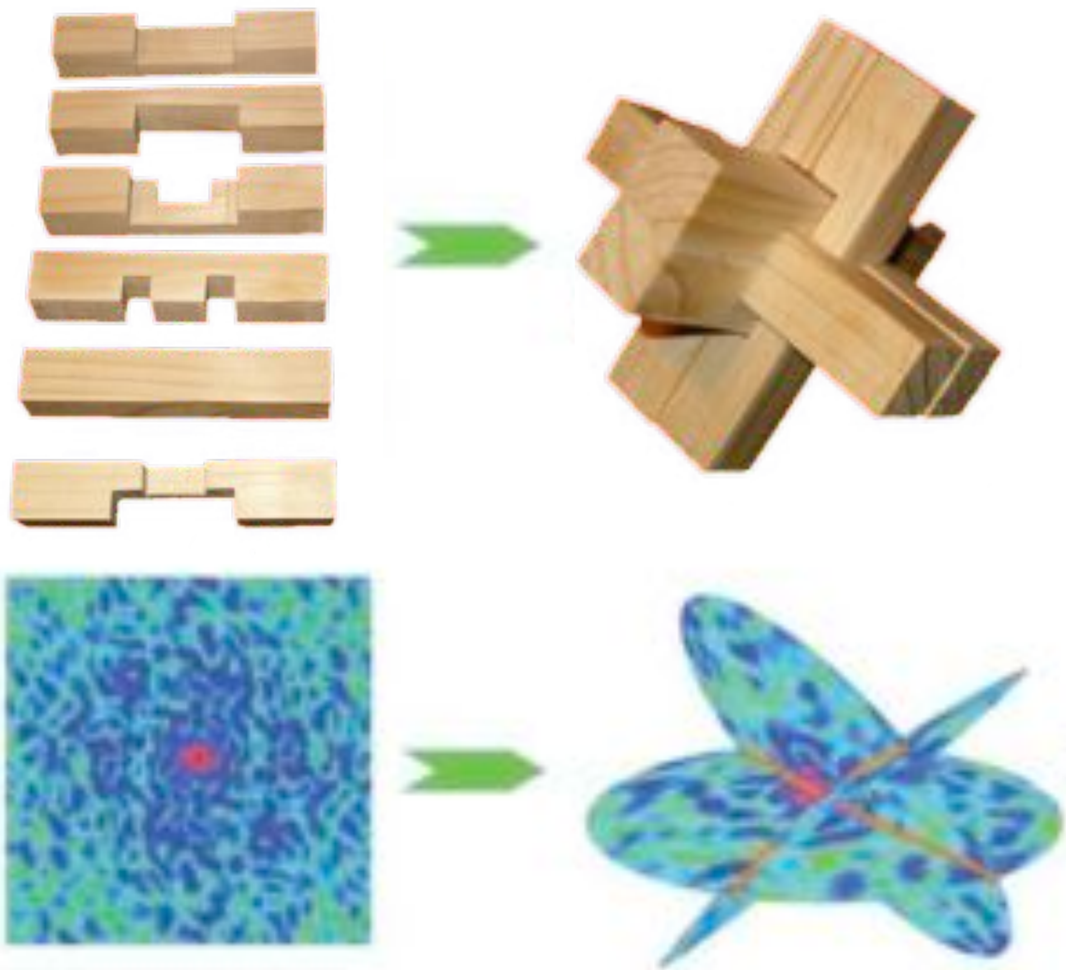
CXIDB
Coherent X-ray Imaging Data Bank

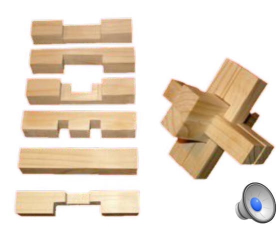
Hawk
Image Reconstruction



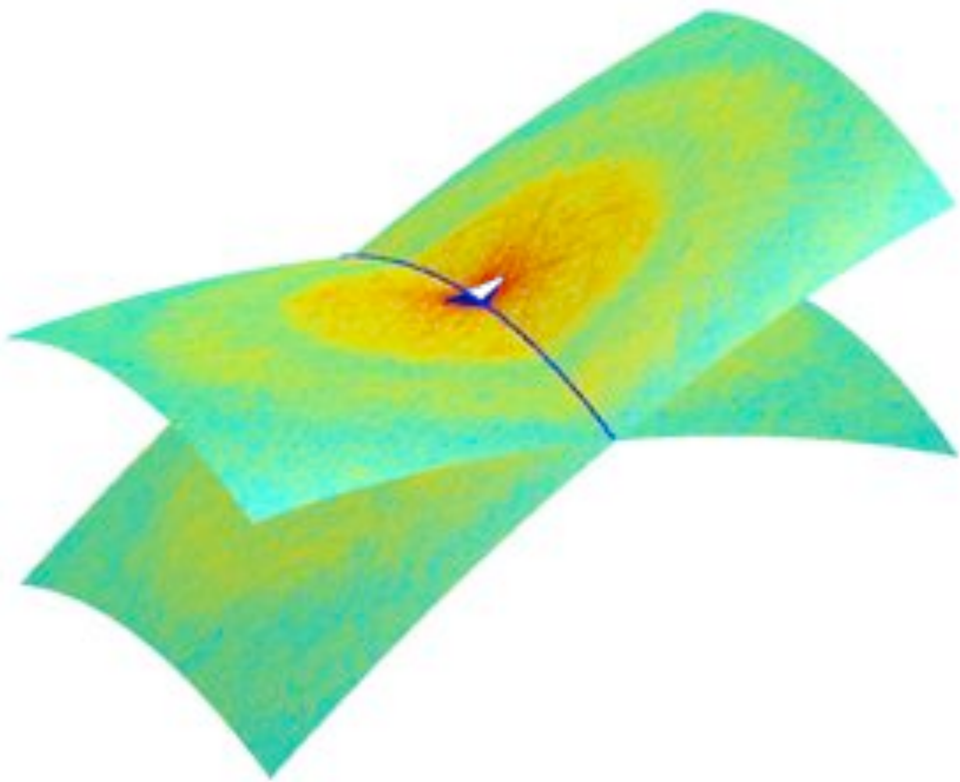
3D Assembly

3D assembly of the diffraction data is a hard puzzle

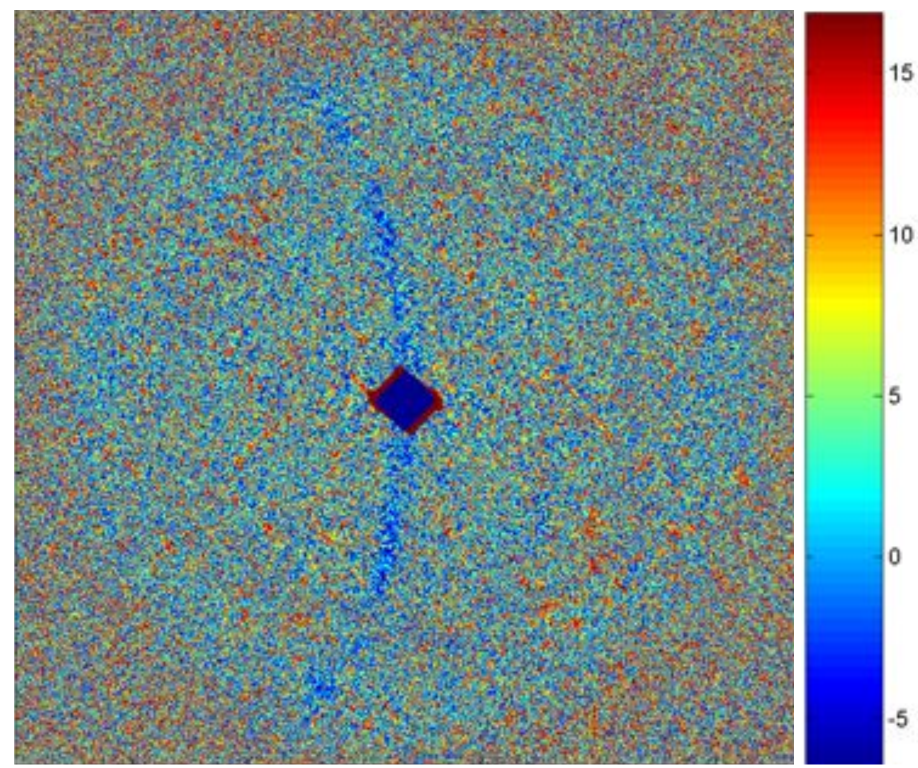




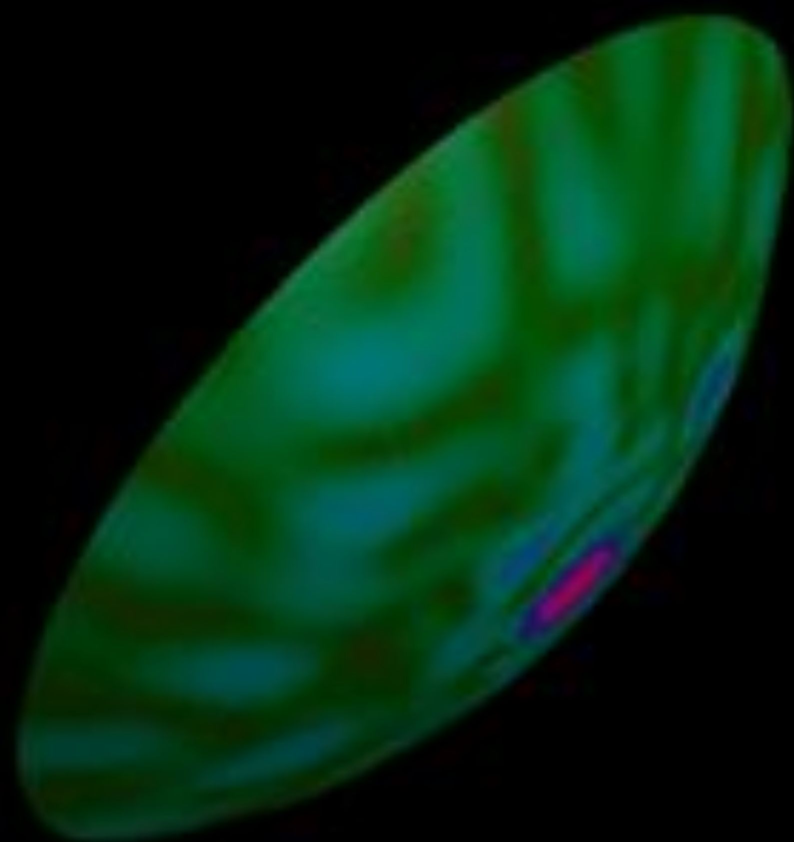
Algorithms developed for electron microscopy can in principle be used to solve the problem



Difference of patterns at two object orientations

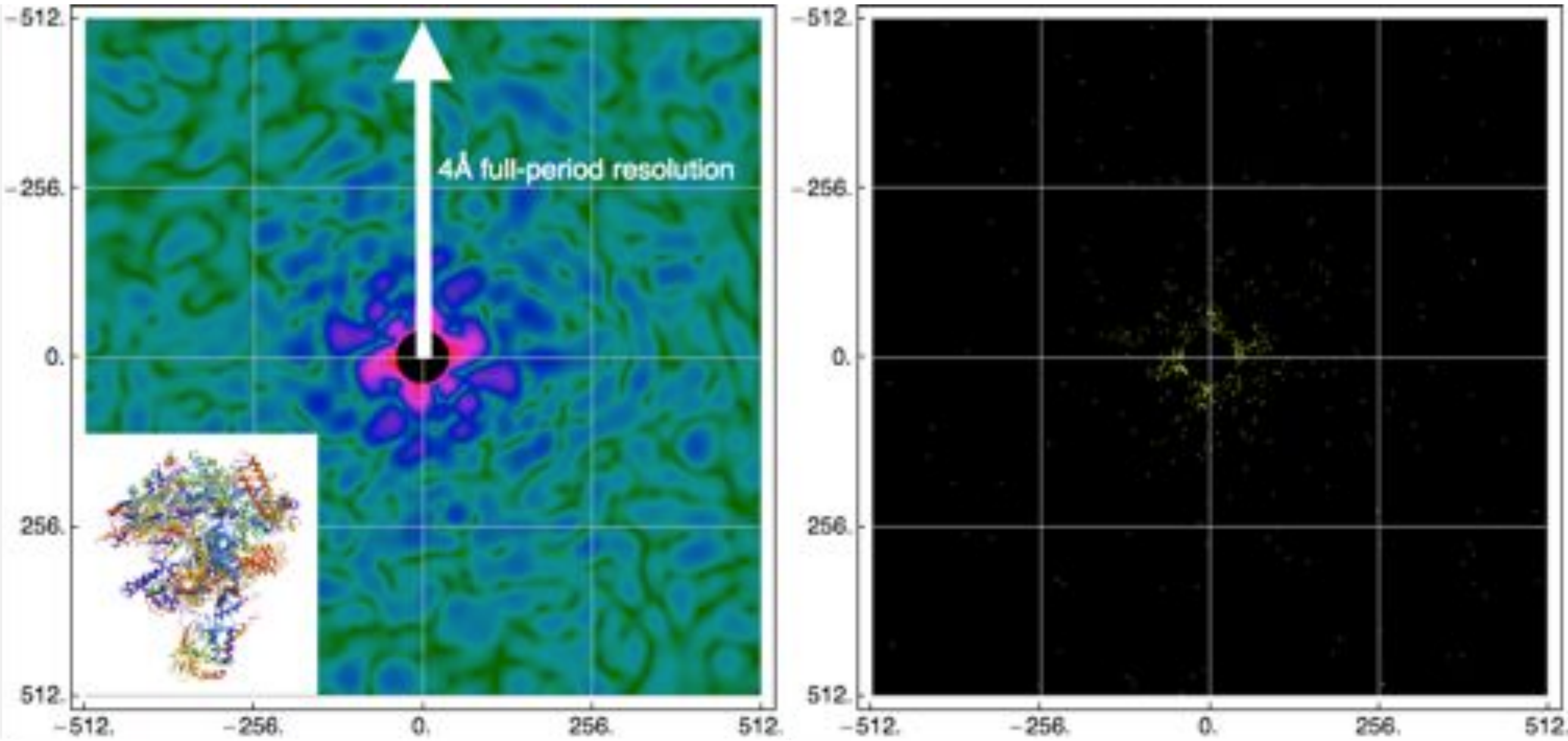


Gösta Huldt *et.al*.
J.Struct.Biol. 144, 219 (2003)



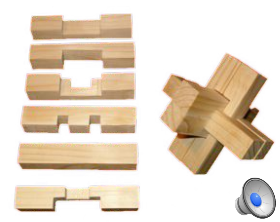
Duane Loh, SLAC

In practice the signal is too weak and new algorithms are required



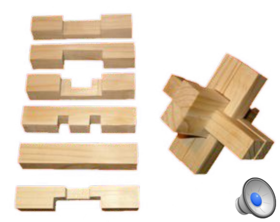
10^12 incident photons per pulse
100nm focus diameter,
1.5Å radiation

~600 photons



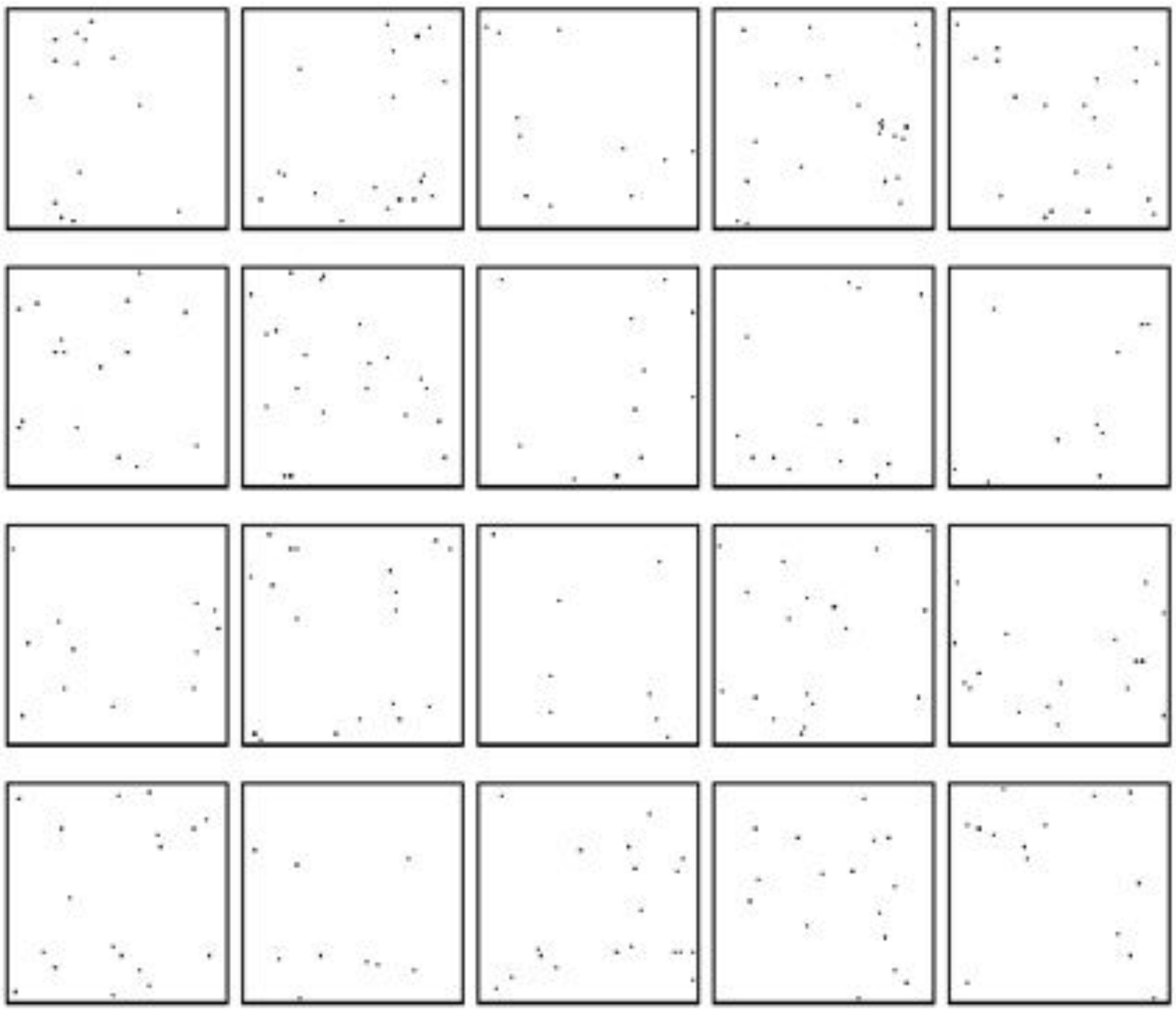
Duane Loh, SLAC

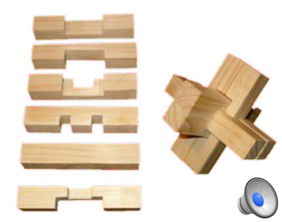
Example of the Expand Maximize Compress (EMC) algorithm in 2D



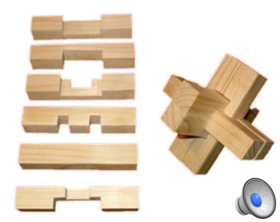
Duane Loh, SLAC

The noisy images

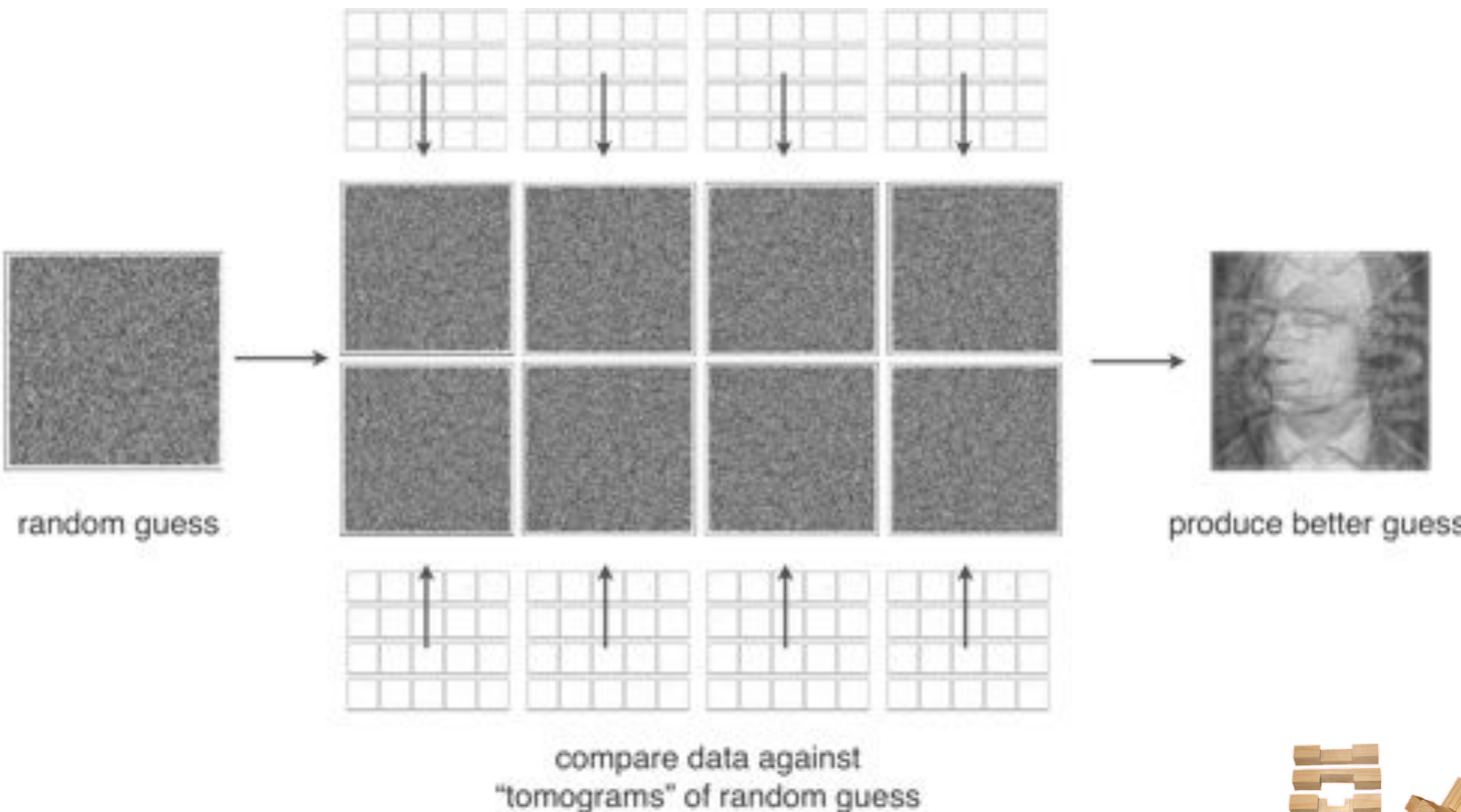


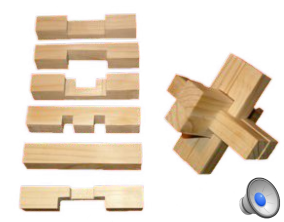


If we knew the solution we could just compare the data against it to determine the orientation

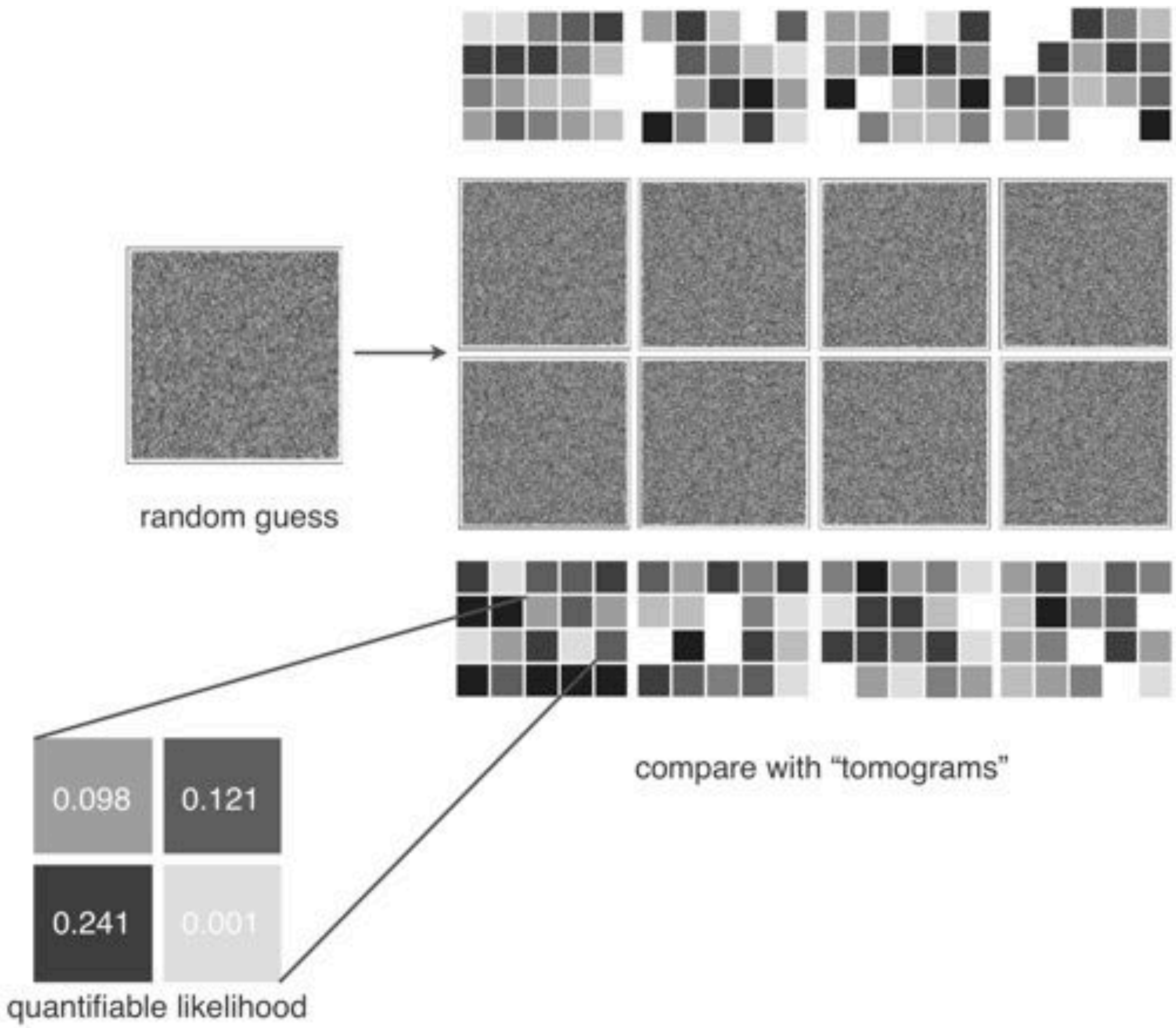


In practice we start from a random model and use the output as the model for the next iteration

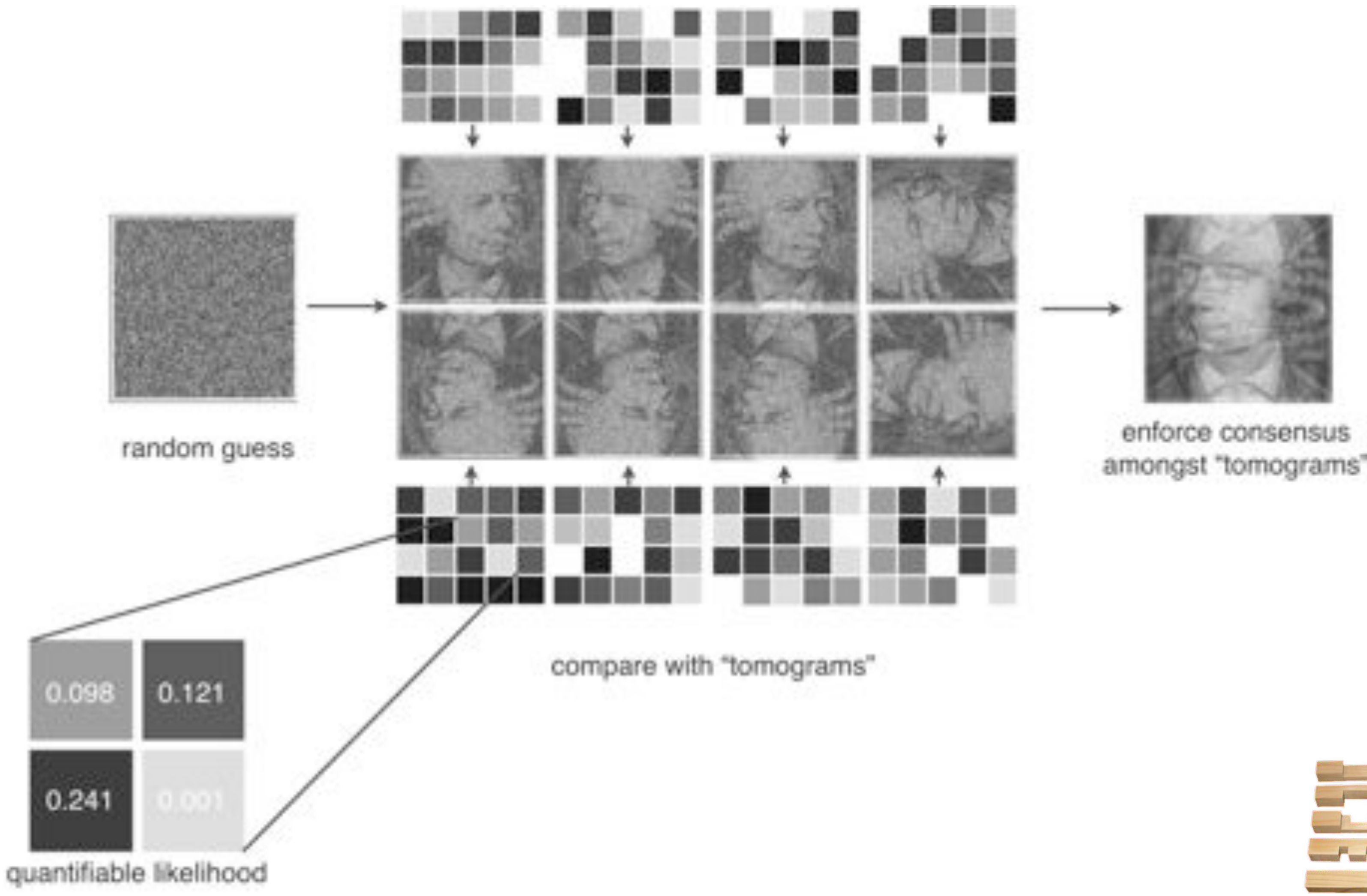


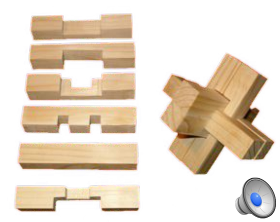


The probability that the images correspond to each orientation of the model is calculated

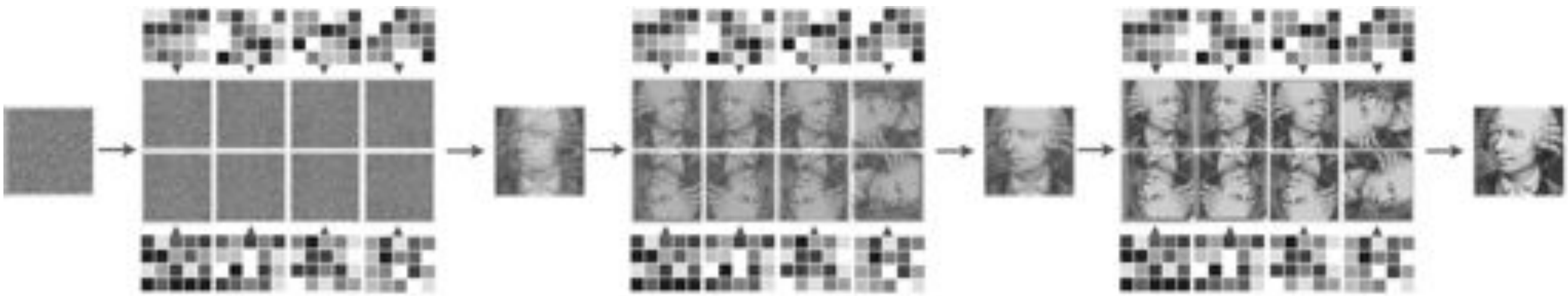


The images, weighted by the probabilities are combined in a new model





The process has to be repeated iteratively until the solution is reached



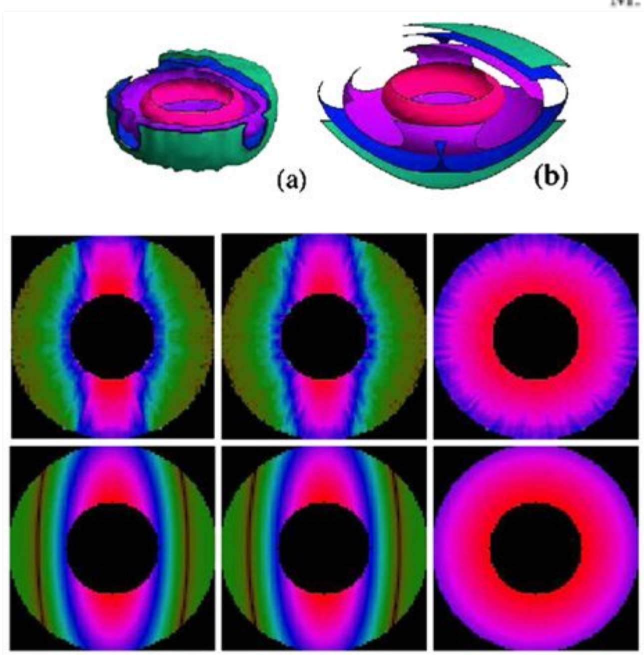
Veit Elser, Cornell University.

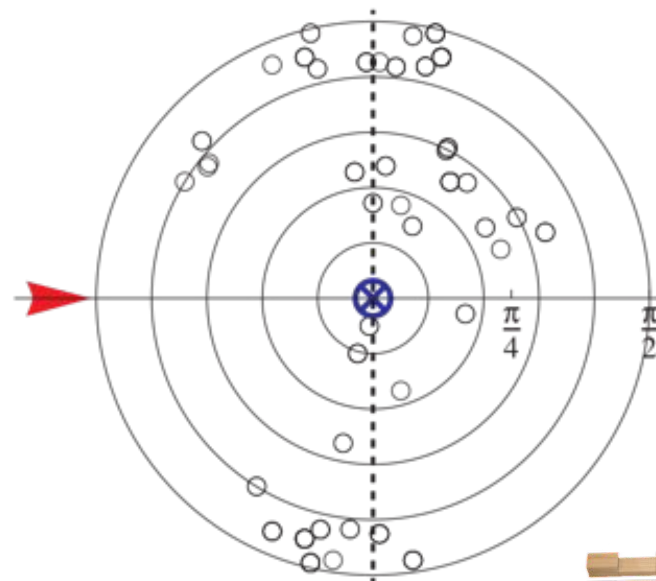
The method was demonstrated experimentally

PRL 104, 225501 (2010) PHYSICAL REVIEW LETTERS week ending 4 JUNE 2010

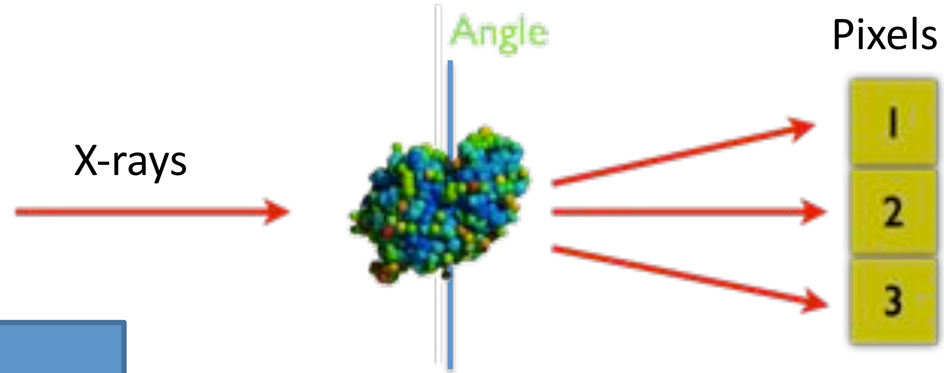
Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns

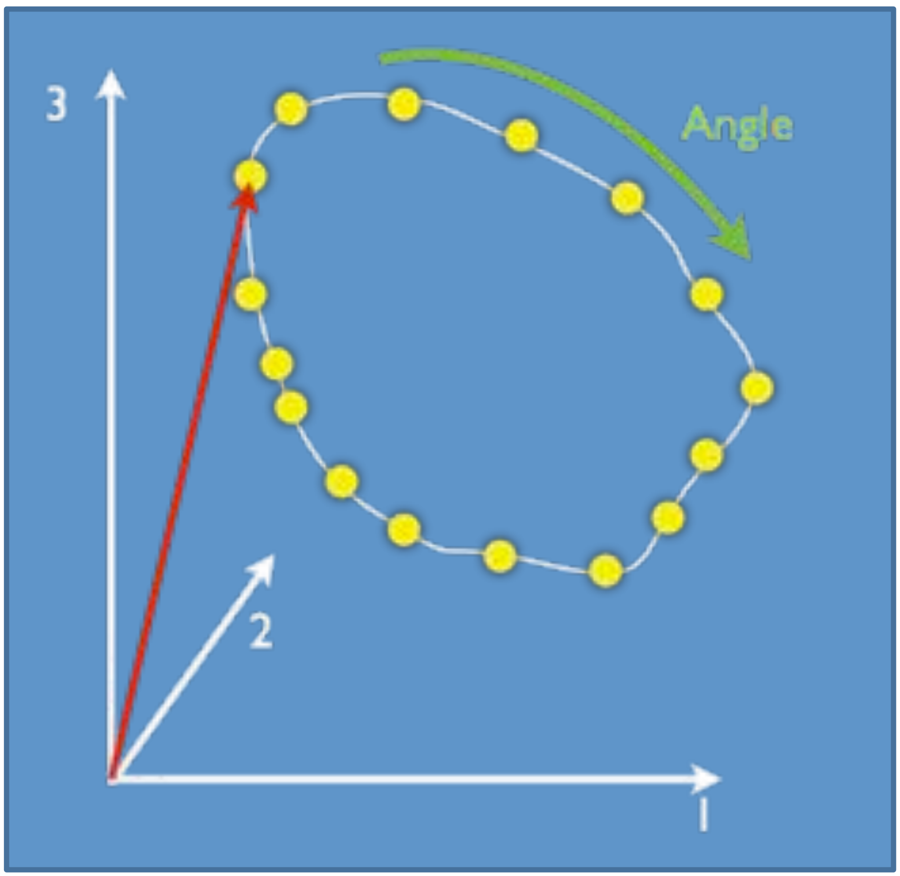
N. D. Loh,[1,11] M. J. Bogan,[2] V. Elser,[1] A. Barty,[3] S. Boutet,[2] S. Bajt,[4] J. Hajdu,[5] T. Ekeberg,[5] F. R. N. C. Maia,[5] J. Schulz,[3] M. M. Seibert,[5] B. Iwan,[5] N. Timneanu,[5] S. Marchesini,[6] I. Schlichting,[7,8] R. L. Shoeman,[7,8] L. Lomb,[7,8] M. Frank,[9] M. Liang,[3] and H. N. Chapman[3,10]





Manifold embedding based algorithms are another possible alternative



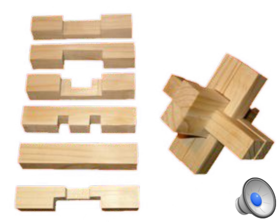


1. Pixels form an n-dimensional hyperspace

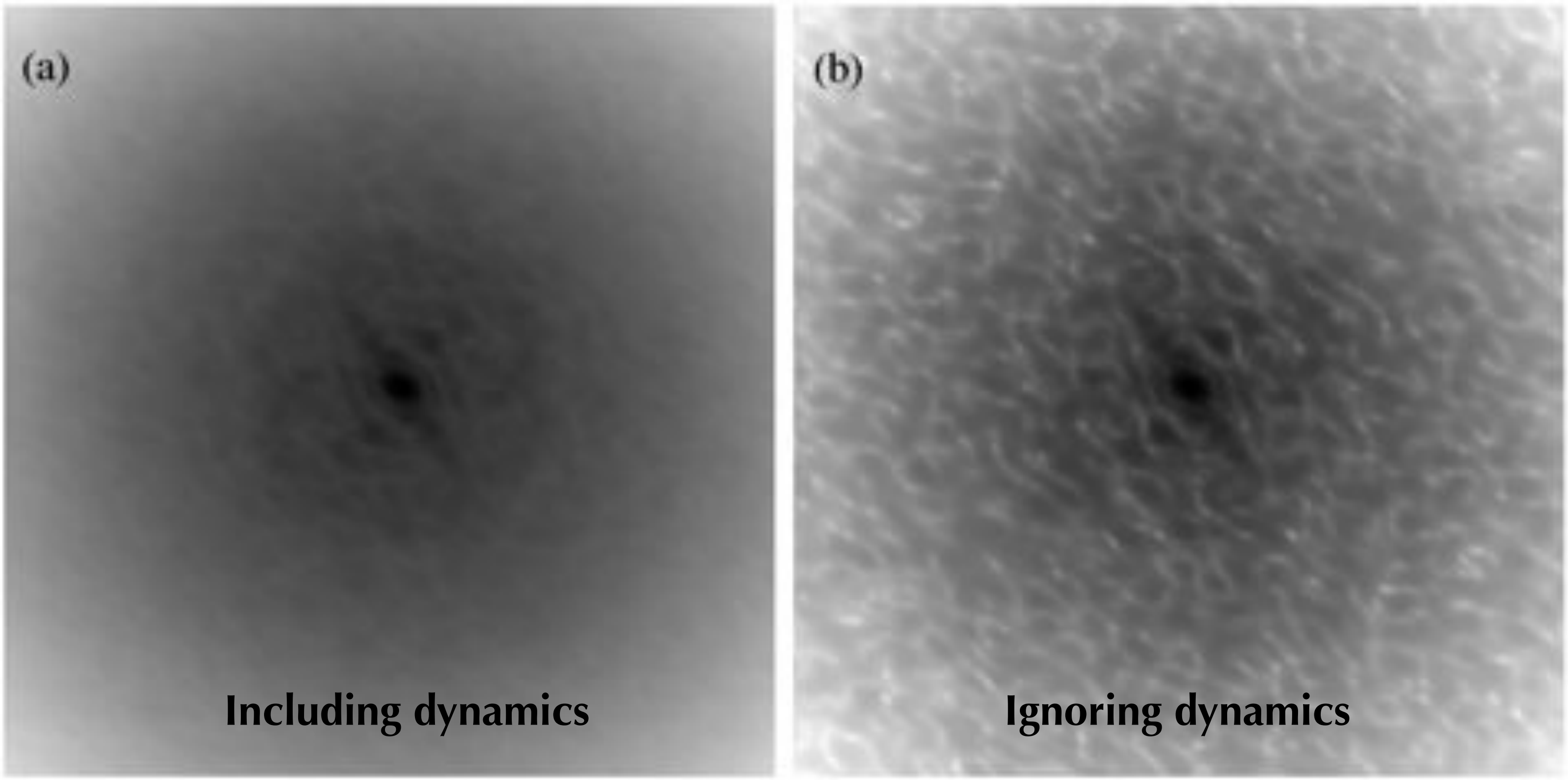
2. Each diffraction pattern is a unique point in n-space

3. Points move in n-space as the sample rotates

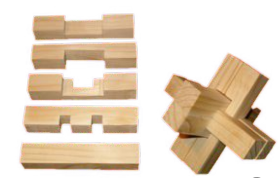
4. Euclidean metric links adjacent points into a smooth manifold encoding sample rotation



Sample variability is another obstacle to high resolution

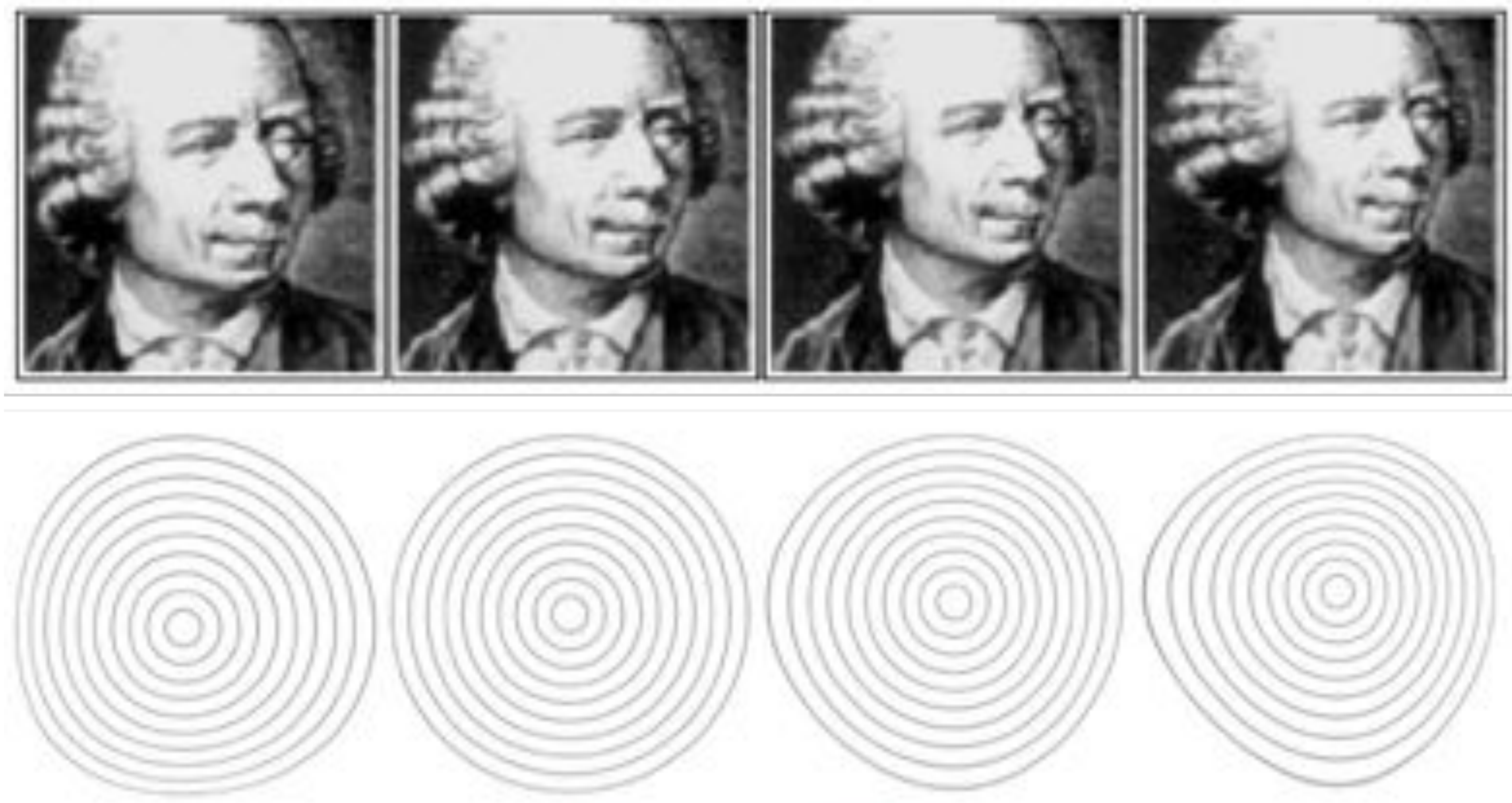


Averaged diffraction patterns from Lysozyme

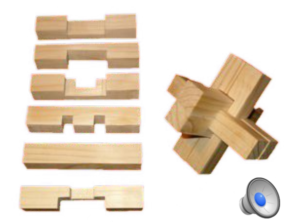


Maia, F.R.N.C., Ekeberg, T., Timneanu, N., van der Spoel, D. & Hajdu, J.
Physical Review E 80, 1-6 (2009).

EMC can be used to distinguish a limited amount of variability

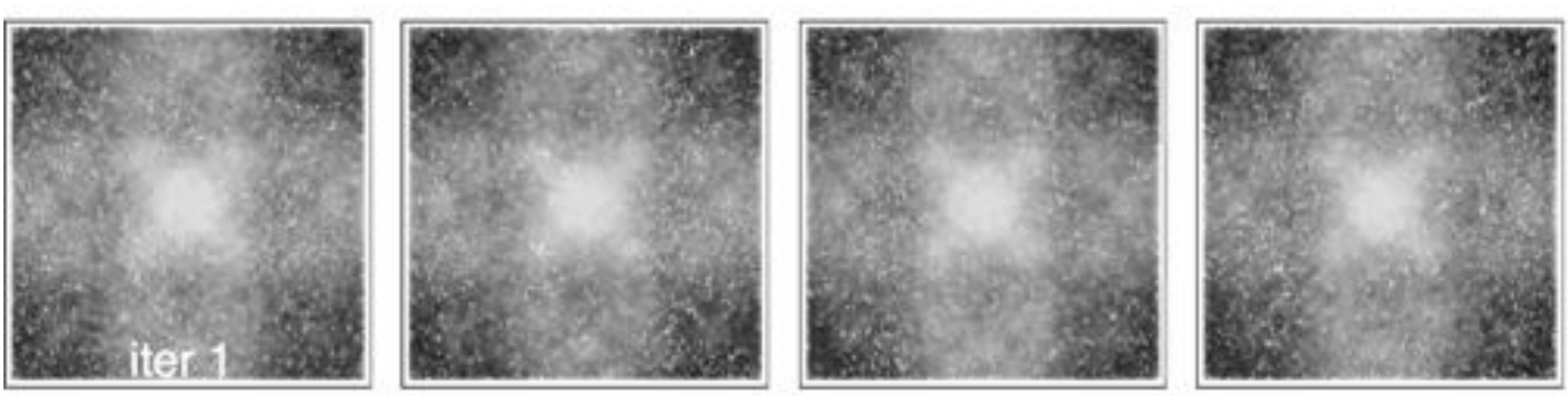


Data includes distortions of
original image.



Duane Loh, SLAC

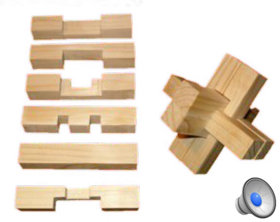
EMC can be used to distinguish a limited amount of variability



EMC can be used to distinguish a limited amount of variability

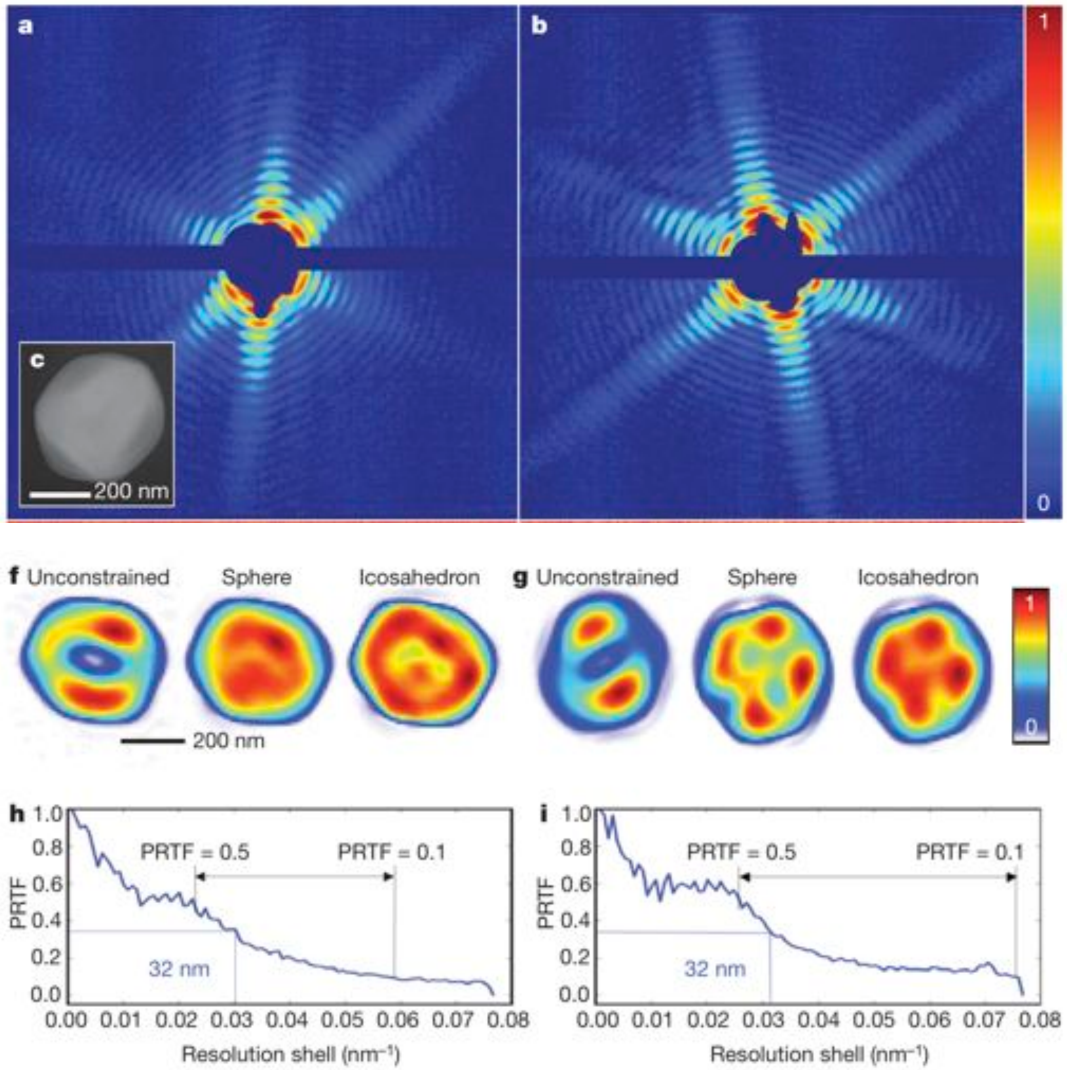
Recovered Images

Solution



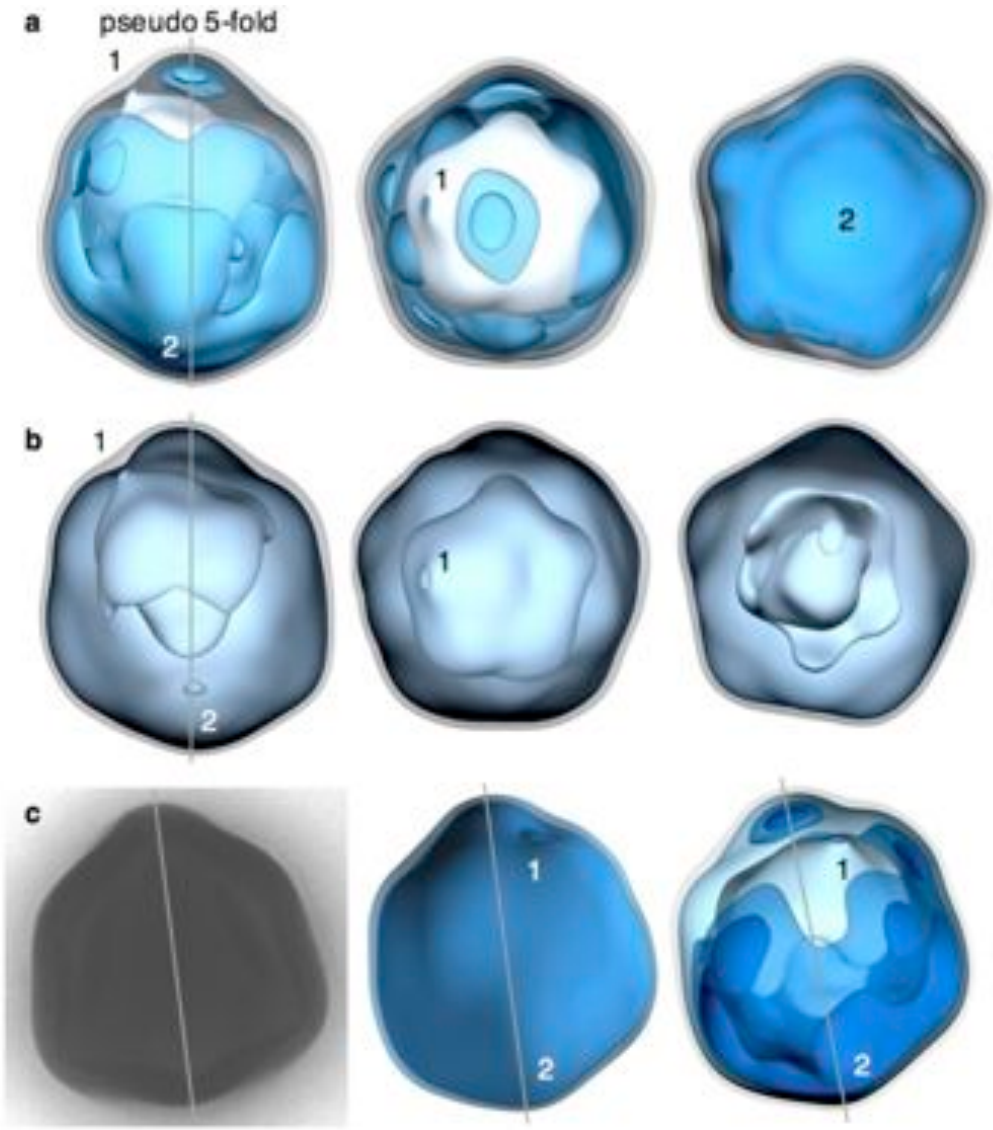
Validation

Validation and correct interpretation of results can be hard due to the complexity of the method



Seibert, M.M. et al.
Nature 470, 78-81 (2011).

The introduction of the 3D assembly brings a whole new set of possible sources of error



Tomas Ekeberg,
Uppsala University
manuscript in preparation

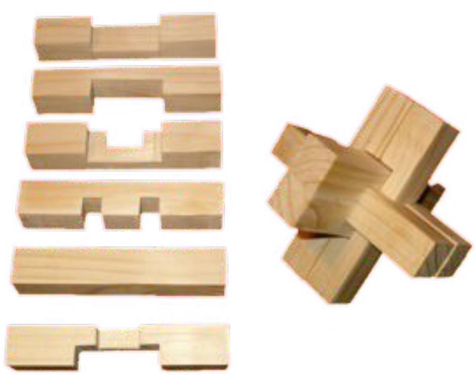
Thank you.

Questions?

Automation

What are the major challenges ahead for X-ray Coherent Diffraction Imaging?

Data Deluge

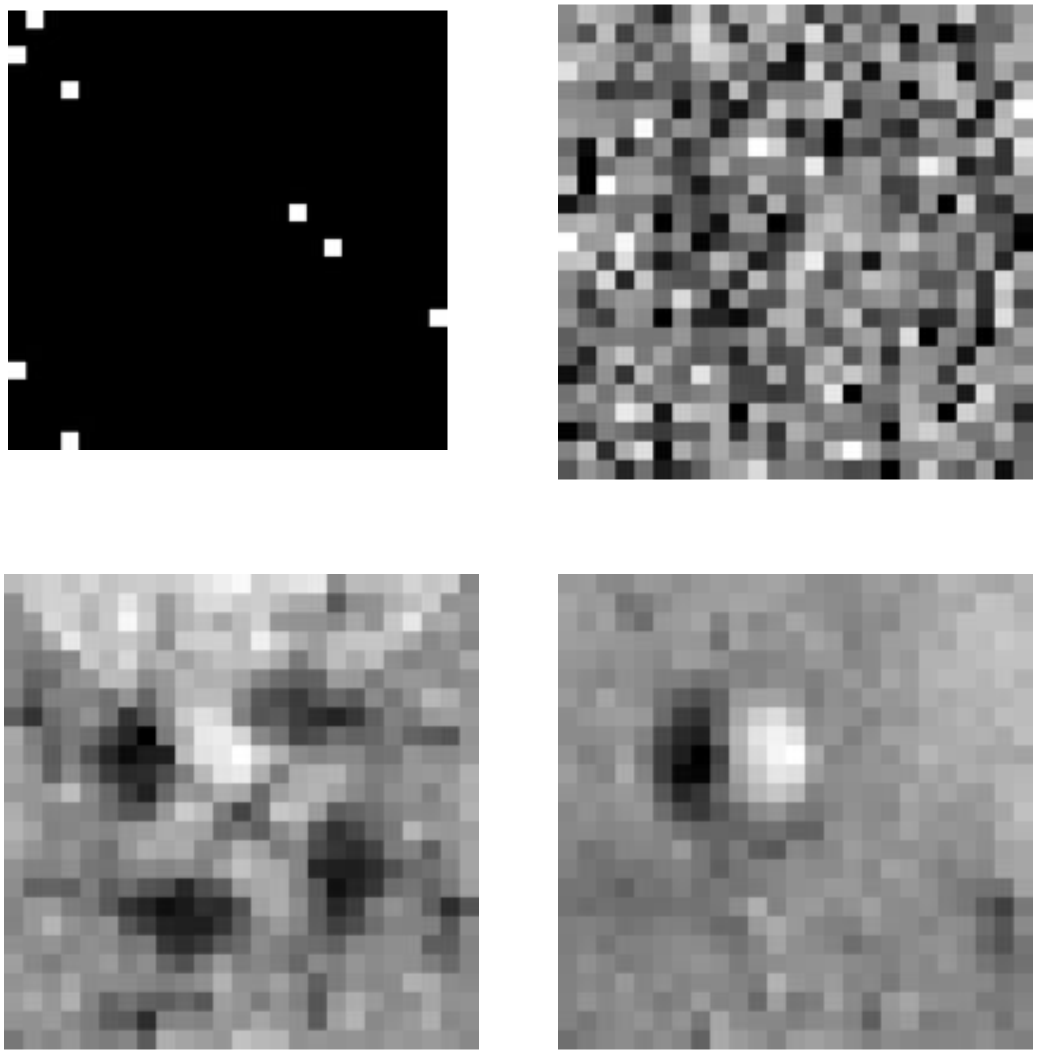


3D Assembly

Validation

Automation

The EMC algorithm fits each image against a model, and combines them to create a new model



1. Start with a random model
2. Compare experimental images to model in all possible positions.
3. Combine the experimental images in a new model according to how well they fit the previous model in each position.
4. Go to 2.